library(ggplot2)

# Loading and inspecting the dataset

data = read.csv('hospital-data.csv')  
#str(data)  
nrow(data)

## [1] 4826

sum(!complete.cases(data)) # Check for missing data the general way

## [1] 4826

# A more robust approach is to iterate over all columns  
output <- vector("double", ncol(data)) # Initialize an output vector  
for(col in seq\_along(data)){  
 output[[col]] <- sum(is.na(data[[col]]))  
}  
output

## [1] 0 0 0 4826 4826 0 0 0 0 0 0 0 0

#From the above it appears there is completely no data in columns 4 & 5  
#We can remove them from our dataset  
data = data[,-(4:5)]  
dim(data)

## [1] 4826 11

head(data) #Peek at your data

## Provider.Number Hospital.Name Address.1  
## 1 10001 SOUTHEAST ALABAMA MEDICAL CENTER 1108 ROSS CLARK CIRCLE  
## 2 10005 MARSHALL MEDICAL CENTER SOUTH 2505 U S HIGHWAY 431 NORTH  
## 3 10006 ELIZA COFFEE MEMORIAL HOSPITAL 205 MARENGO STREET  
## 4 10007 MIZELL MEMORIAL HOSPITAL 702 N MAIN ST  
## 5 10008 CRENSHAW COMMUNITY HOSPITAL 101 HOSPITAL CIRCLE  
## 6 10009 HARTSELLE MEDICAL CENTER 201 PINE STREET NORTHWEST  
## City State ZIP.Code County Phone.Number Hospital.Type  
## 1 DOTHAN AL 36301 HOUSTON 3347938701 Acute Care Hospitals  
## 2 BOAZ AL 35957 MARSHALL 2565938310 Acute Care Hospitals  
## 3 FLORENCE AL 35631 LAUDERDALE 2567688400 Acute Care Hospitals  
## 4 OPP AL 36467 COVINGTON 3344933541 Acute Care Hospitals  
## 5 LUVERNE AL 36049 CRENSHAW 3343353374 Acute Care Hospitals  
## 6 HARTSELLE AL 35640 MORGAN 2567736511 Acute Care Hospitals  
## Hospital.Ownership Emergency.Services  
## 1 Government - Hospital District or Authority Yes  
## 2 Government - Hospital District or Authority Yes  
## 3 Government - Hospital District or Authority Yes  
## 4 Voluntary non-profit - Private Yes  
## 5 Proprietary Yes  
## 6 Proprietary Not Available

summary(data) # Summary statistics

## Provider.Number Hospital.Name Address.1 City   
## Length:4826 Length:4826 Length:4826 Length:4826   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## State ZIP.Code County Phone.Number   
## Length:4826 Min. : 603 Length:4826 Min. :9.369e+08   
## Class :character 1st Qu.:32208 Class :character 1st Qu.:3.616e+09   
## Mode :character Median :54154 Mode :character Median :6.033e+09   
## Mean :53079 Mean :5.829e+09   
## 3rd Qu.:75226 3rd Qu.:7.877e+09   
## Max. :99929 Max. :9.899e+09   
## Hospital.Type Hospital.Ownership Emergency.Services  
## Length:4826 Length:4826 Length:4826   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

sapply(data, mode)

## Provider.Number Hospital.Name Address.1 City   
## "character" "character" "character" "character"   
## State ZIP.Code County Phone.Number   
## "character" "numeric" "character" "numeric"   
## Hospital.Type Hospital.Ownership Emergency.Services   
## "character" "character" "character"

# Exploratory Data Analysis

Depending on the mode/data type of each variable, you will perform specific EDA technique.

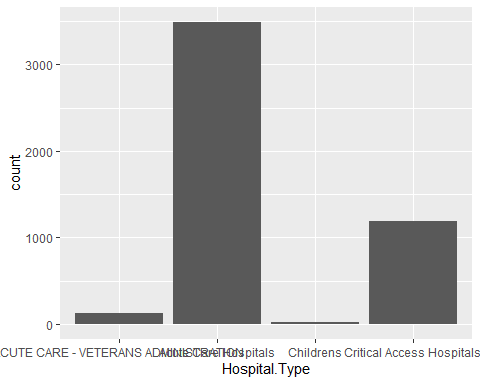
# Categorical variables  
data$Hospital.Type = factor(data$Hospital.Type)  
data$Emergency.Services = factor(data$Emergency.Services)  
table(data$Emergency.Services)

##   
## No Not Available Yes   
## 231 23 4572

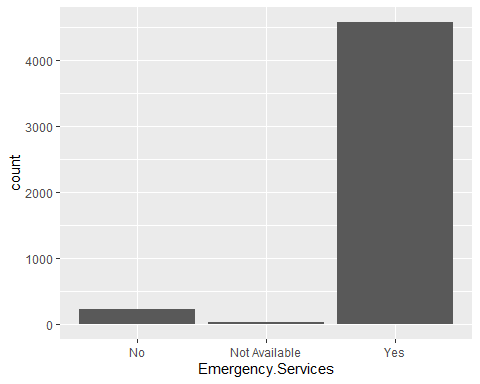
table(data$Hospital.Type)

##   
## ACUTE CARE - VETERANS ADMINISTRATION Acute Care Hospitals   
## 129 3491   
## Childrens Critical Access Hospitals   
## 22 1184

#Graphics  
ggplot(data = data) +  
 geom\_bar(aes(x = Hospital.Type))



ggplot(data = data) +  
 geom\_bar(aes(x = Emergency.Services ))



# Continuous/numeric columns, you can try scatter plots for bivariate data and histograms and density/frequency polygons plots for univariate data